Development of a Bayesian Network for Diagnosis of Breast Cancer

Linda M. Roberts, Charles E. Kahn, Jr., Peter Haddawy
{roberts,kahn,haddawy}@cs.uwm.edu

Department of Electrical Engineering and Computer Science
University of Wisconsin-Milwaukee
P.O. Box 784
Milwaukee, Wisconsin 53201 USA

Department of Radiology
Medical College of Wisconsin
8700 West Wisconsin Avenue, DH Box 151
Milwaukee, Wisconsin 53226 USA

Dr. Haddawy: Telephone (414) 229-4955 • Fax (414) 229-6958

We describe the early stages of the development and validation of a Bayesian network to assist in the detection of breast cancer. MammoNet integrates mammographic findings, demographic factors, and physical examination to determine the probability of malignancy. Conditional probabilities were obtained from the medical literature and from expert opinion. Problems (and solutions) encountered while developing the model are discussed. MammoNet is implemented as a knowledge base of rules; problem-specific networks are constructed using a Bayesian network construction algorithm. The model was validated by evaluating its performance on a suite of textbook cases.

Keywords: breast cancer, mammography, model structure, probability assessment, probabilistic rules, BNG

Submitted to the IJCAI-95 Workshop on Building Probabilistic Networks

1 Introduction

Breast cancer accounts for approximately 40,000 deaths of American women annually; 120,000 to 180,000 new cases of the disease are detected each year [9]. A woman has a lifetime risk of 1 in 9 of developing breast cancer [23]. Screening mammography is an effective method for detecting early breast cancers in asymptomatic women, which increases the likelihood for cure and long-term survival [20, 16]. A suspicious or indeterminate mammographic finding can lead the attending physician to perform a breast biopsy. But as many as 75% of certain mammograms fall into indeterminate grades of radiological suspicion [1]. One would like to avoid recommending unnecessary biopsy because of the costs and discomfort to the patient and because biopsy can confound later mammographic testing by producing radiographic abnormalities which can be mistaken for cancer [21]. Using biopsies in excess for benign conditions also increases the costs of mammographic screening and represents a barrier to the effective utilization of this resource [2]. Thus we were motivated to develop a decision support tool to aid in the evaluation of mammographic findings.

This paper describes the design, implementation, and preliminary evaluation of a Bayesian network that models the presence or absence of breast cancer based on demographic risk factors, mammographic signs, and physical findings. The network is implemented using the Bayesian Network Generation system (BNG) [11], which generates Bayesian networks from knowledge bases consisting of first-order probability logic sentences. Given a knowledge base of rules, a set of evidence E, and a query Q, the BNG system generates the structurally minimal network to compute P(Q | E). (A network is structurally minimal if it contains no barren nodes and no nodes d-separated from the query.) The generated network is passed to the IDEAL system [22] to perform the probabilistic computation.

2 Problem Overview

Mammography is an important tool in early detection of breast cancer. Indeterminate mammographic findings often challenge the physician to distinguish between a malignant or benign condition. Successful diagnosis depends on the ability of a physician to recognize and evaluate a mammographic abnormality, as well as integrate the information from multiple clinical aspects (e.g., risk factors, physical findings) to determine the likelihood of breast cancer [9].

2.1 Risk Factors and Physical Symptoms

Recent demographic investigations as described by Gail et al. [10] and Colditz [4] have reported several risk factors considered to increase a woman's chance of developing breast cancer. Breast cancer incidence rates increase with age. Breast cancer is uncommon for a woman less than age 40, though this group shows the fastest rate of increase of the disease. Positive diagnoses occur with greater frequency in post-menopausal women. An early menarche, with its associated early onset of regular menstrual cycles, is another accepted risk factor. A late child bearing age, that is giving birth after age 30 or never giving birth, raises the risk. Women with a first-degree relative who has had breast cancer double their chances of developing breast cancer. Symmonds [23] found that breast pain, nipple discharge, and skin thickening are reported by women with breast cancer, but few early-stage cancers are detected by these indicators.

2.2 Mammographic Findings

None of the established risk factors practically translate into preventive measures, making mammography an important screening tool. Mammographers group breast cancer findings into direct and indirect signs. Direct mammographic indications are calcified or noncalcified masses or calcifications alone. Indirect mammographic signs include architectural distortion, asymmetry, dilated duct, and developing density.

Benign and malignant masses are differentiated through mass attributes of margin, density, location, and the presence of the halo sign [24]. Round, low-density masses with smooth, sharply defined margins are considered benign. High-density, stellate, spiculated or knobby masses with poorly defined margins are considered malignant [19]. Frequently, though, masses are classified as indeterminate, not clearly benign or malignant. Improved mammography and increased screening of asymptomatic women have increased the number of cancerous masses detected not displaying the characteristic radiographic features of malignancy. Instead of spiculated masses, many masses display as nondescript lesions having poorly defined or irregular margins [19].

Similarly, the attributes of size, shape, density, distribution pattern, and number are examined when differentiating between benign and malignant calcifications. Benign and malignant calcifications can occur with or without a mass. Benign calcifications are typically large (1-4 mm in diameter), coarse, round or oval, and monomorphic (uniform in size and shape). Their distribution pattern is typically scattered or diffuse. If the calcifications are clustered, they number less than 5 per cluster. Some benign calcifications display bizarre, irregular shapes, but because of their large size are considered noncancerous [19, 2, 18]. Malignant calcifications are typically

microscopic (<0.5 mm in diameter) and fine, linear branching or rod-shaped, punctate- or stellate-shaped, and pleomorphic (varying in size and shape). Their distribution pattern is grouped or clustered, and they are innumerable. A rule of thumb is the greater the number of calcifications in a cluster (usually greater than 5), the greater the likelihood of malignancy [19, 2, 18]. As with breast masses, calcifications can display indistinct characteristics making the determination of malignancy difficult. Both benign and malignant calcifications can appear tiny and clustered [19]. Typically, malignant calcifications present with a wide range in size, shape, and density [8].

Sickles [20] notes that almost 20% of cancers present with neither mass nor calcifications, but with subtle or "indirect" signs of malignancy. Architectural distortion is a frequent indicator of breast cancer. However, a woman who has undergone a breast biopsy can present with this indication. Developing density, an enlarging area of glandular tissue density, is a strong indicator of cancer. Dilated ducts and breast asymmetry (increased density as compared with mirror-image location in the opposite breast) are less effective indicators.

3 Model Structure

MammoNet is a medical decision-support system to aid in diagnosing breast cancer given demographic risk factors, radiological information, and physical symptoms. The first task in the modeling of MammoNet was identifying the hypothesis. Our hypothesis events, *breast cancer present* and *breast cancer absent*, were organized into one node *BreastCancer* with states present and absent.

Following the standard classification of information related to breast cancer, we divided the random variables in our domain into three logical classes: demographic features, mammographic indications, and physical findings. Demographic features consist of variables for the established risk factors, age, age of menarche, age at first live birth, and number of first degree relatives with breast cancer. Mammographic indications are subdivided into direct and indirect indications. Variables in this class model the direct signs masses, calcifications, and their characteristics, and the indirect signs, architectural distortion, asymmetry, and developing density. Variables in the third class model physical symptoms, pain and nipple discharge.

Our next step in the model building was establishing the causal relations between variables. Demographic features cause breast cancer. Breast cancer causes a calcified or noncalcified mass or calcifications only. A mass or calcification will possess certain characteristics depending on whether it is malignant or benign. Breast cancer causes architectural distortion, asymmetry, and developing density. Breast cancer causes pain and nipple discharge .

A problem we grappled with during the design of MammoNet was the relationship between calcified or noncalcified masses and calcifications alone. How are masses and calcifications related? The literature was contradictory. Sickles [20] analyzed masses and calcifications separately; Symmonds [23] all three. An early design modeled these findings as three nodes having states, present and absent. A finding's characteristic (mass margin) was grouped with its parent node (mass). This design did not accurately propagate the degree of belief in breast cancer conditioned on different characteristics. A second attempt to capture this relationship used a "mammographic findings" node having states, mass, mass with calcifications, calcifications, and none. We lost the concept of malignancy in this model. A third network returned to three nodes but with states, malignant, benign, and none. A constraint node embodied the logical dependence between these nodes. A consultation with a mammographic expert from the Medical College of Wisconsin resolved this issue and simplified our model. The expert mammographer, when presented with a calcified mass, reasons about each finding independently. She estimates a cancer probability based on the mass's characteristics, and then she estimates a probability based on the calcification's characteristics. She combines both estimates for a final estimate for the calcified mass. MammoNet's final structure contains nodes for mass and calcification having states, benign, malignant, and none.

The mammographic expert also explained the relation between the indirect sign, architectural distortion and a previous biopsy at the same site. Architectural distortion without a previous biopsy at the same site is "an automatic request for a biopsy" according to our expert. However, a previous biopsy at the same site is a reasonable explanation for an architectural distortion. Both breast cancer and previous biopsy at the same site are causes of architectural distortion. Positive evidence for a previous biopsy at the same site explains away breast cancer as a cause of architectural distortion.

4 Data Acquisition

Statistical studies published in radiology journals provided most of the data for MammoNet's knowledge base. When required probability data were unavailable or the sample size too small, we consulted the mammographic expert. She provided subjective estimates of the probabilities for architectural distortion, previous biopsy at the same site, and the halo sign.

Prior probabilities are assigned to the demographic nodes. Age statistics were obtained from the National Cancer Institute and the U.S. Census Bureau, Population Division, release PPL-8. Statistics for age of menarche were acquired from the Department of Health, Education, and Welfare, Vital and Health Statistics. Population and Vital Statistics, Statistical Record of Women

Worldwide provided statistics for age of first live birth. The remaining demographic feature, number of first degree relatives with a known history of breast cancer, was estimated based on information from Colditz [4].

Conditional probabilities for the physical findings given breast cancer were obtained from Symmonds et al. [23], for the mammographic findings given breast cancer from Sickles [19, 20], Egan et al. [8], and Jackson et al. [15], and for the indirect mammographic findings given breast cancer from Sickles [20].

We used a generalized noisy-or gate [5] to model the multi-causal influence on the node *BreastCancer*. *BreastCancer* is influenced by the four demographic risk factors: age, age of menarche, age of first live birth, and number of first degree relatives with a history of breast cancer. The risk factors are independent and mutually reinforcing in the sense that the more factors present, the more likely *BreastCancer* will be in state present. On the average, a nulliparous post-menopausal woman, whose age of menarche was early, and who has a family history of breast cancer is more likely to develop breast cancer than a pre-menopausal woman who has had children at an early age, whose age of menarche is late, and who has no family history of breast cancer.

The link matrix contains 12 * 3 * 4 * 3 = 432 entries, all combinations of the states of the nodes. Using data retrieved from the National Cancer Institute's Surveillance Program, we determined the probability of developing breast cancer in a specific time interval. We multiplied the Relative Risk(RR) factors computed in [10] by the base risk factor for a specific age group to obtain an adjusted probability. A 50-year old woman with no history of breast cancer runs a 2 % base line risk of developing cancer. An age of menarche in the 12-14 year range, adds a RR factor of 1.099. Similarly, a first birth in the range of 20-24 years, and no history of breast cancer in first degree relatives, adds a RR factor of 1.244. The adjusted probability for developing breast cancer is then computed by multiplying the base factor by a product of the additional risk factors.

Sample computation of risk factors for breast cancer:

```
(BASE_RISK) * (RELATIVE_RISK)  = ADJUSTED_RISK
   (0.02)   * (1.099 * 1.244 ) =  0.02734312
```

5 Implementation

The BNG system generates Bayesian problem-specific networks from probability logic knowledge bases. A class of networks is specified by a set of rules representing universally quantified conditional probability sentences. Given a ground query and a set of ground atomic evidence

formulas, BNG generates the structurally minimal network to compute the posterior probability of the query. The network is then passed to the IDEAL system [22] for evaluation. The rules in the knowledge base can be thought of as representing network fragments. For example, the network fragment depicting the direct parent/child relation of the direct mammographic indication that a benign or malignant calcification influences its density is:

```
:P(CalcificationDensity | Calcification)
(    (Calcification ?x (?t))   )
     (;   p(1-2    1-3    2-3    3-4     NA   |  Calcification)
            0.18   0.76   0.04   0.02    0.0  ;  malignant
            0.12   0.48   0.26   0.14    0.0  ;  benign
            0.0    0.0    0.0    0.0     1.0  ;  none
     )
     (1-2  1-3  2-3  3-4  NA)
)
```

The probability that a calcification has a density in the category (2-3) given the calcification is benign is 0.26. We have a "none" state for *Calcification* to capture the knowledge that a calcification is only one possible symptom of breast cancer. Similarly, we use the "none" state in node *Mass*.

Suppose we have the evidence

```
{ Age = 50-54, AgeofMenarch <12, AgeofFLB >=30, NumRel =2,
   TumorLoctnBreast = UO, MassMargin = Spiculated,
   MassDensity = HighDensity }
```

and we wish to know the posterior probability of *BreastCancer*. The network shown in Figure 1 is generated by backward chaining on the query and the given evidence. Notice that the network contains only nodes relevant to the requested computation.

6. Verification and Validation

To test the ability of MammoNet to detect breast cancer, we ran the network against text book cases from Tabar's *Teaching Atlas of Mammography* [25]. We selected our cases to exercise the widest range of states in our network. For each case, Tabar details the mammographic findings, an expert radiologist's diagnosis, and the histological diagnosis based on clinical follow-up and/or biopsy results. We matched the findings to our nodes and states, entered them as evidence, and compared the diagnoses from MammoNet with the radiologist and histological diagnoses. Example test cases are shown in Figure 2.

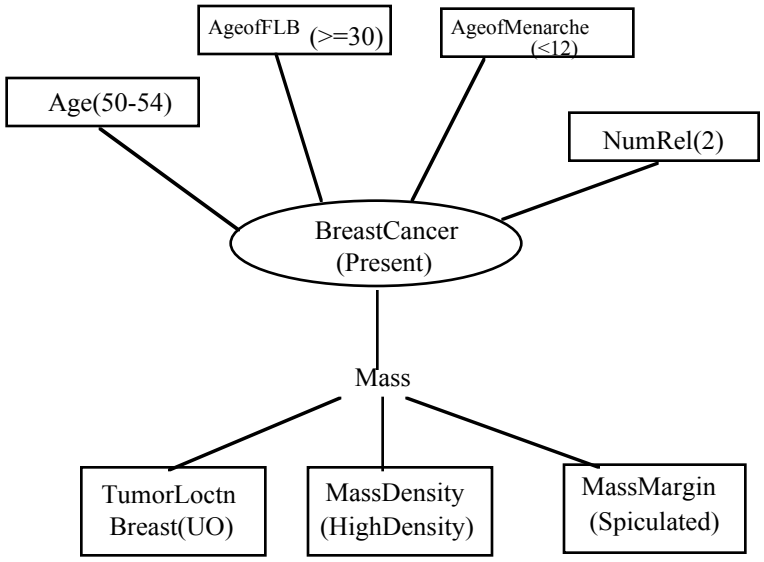

Figure 1: Network generated for query *BreastCancer*(Present)

Age	Mammographic Analysis	Radiologist's Analysis	Histology	MammoNet Coding	MammoNet Result
65	no halo sign, circumscribed, irregular, low density	although circumscribed tumor and low-density, it is irregular and there is no halo sign	mucinous carcinoma; no axillary metastases	MassMargin = irregular; HaloSign = absent; MassDensity = low; Calcification = none	.9960
40	small group of calcification in LO quadrant; casting-type; irregular, varying	cancer	invasive ductal carcinoma; no lymph node metastases	TumorLctnBreast = UO; CalcArrangement = scattered&clustered; CalcShape = LinearBranching	.9659
52	small tumor, UO quadrant, circumscribed, sharply outlined, low density	probably benign	benign	MassMargin = rwdefined; TumorLctnBreast = UO; MassDensity = low	.0020

Figure 2: MammoNet Test Cases

While diagnosing breast cancer based on mass evidence produced highly accurate results, diagnosing based on calcification evidence produced poorer results. We attribute the poorer performance using calcification evidence to the fact that terminology for calcification characteristics is not well defined. *CalcificationDensity*, a node in MammoNet, has the states: 1-2, 1-3, 2-3, 3-4,

and NA. Tabar describes calcification density as "varying." We mapped our state 1-3 to Tabar's "varying," reasoning that the state 1-3 ranges the most from finest to coarsest density.

The performance of the model was assessed using Receiver Operating Characteristic (ROC) analysis [17]. MammoNet's performance — as measured by the very high A_Z value of 0.9585 — compares very favorably with that of artificial neural network models [26] and expert mammographers.

7 Discussion

MammoNet is a Bayesian network designed to assist radiologists in diagnosing breast cancer. It uses the Bayesian Network Generator software package, producing a minimal network. MammoNet successfully models the information to diagnose breast cancer with encouraging preliminary results.

Few problems were encountered during the construction of MammoNet. However, two problems were significant. The scarcity of probabilities for symptoms given breast cancer was troublesome during the design phase. Much of our statistical data was extracted from a small set of articles. There are numerous types of breast cancer, but we were unable to find statistics that related symptoms to any particular type of breast cancer. The literature [9] indicated that different breast cancers present with different symptoms. MammoNet's *BreastCancer* node is generic.

The nonstandard terminology used in the literature was problematic. Where applicable we used Breast Imaging Reporting and Data System (BIRADS) terminology in our model. BIRADS is a standardization of breast imaging terminology for use in medical databases. The nonstandard terminology created problems in both the design stage and testing stage. Calcification terminology was ambiguous, and terms were interchangeable. Malignant calcifications often display a linear branching geometric pattern. Linear branching and casting are both terms used to describe this branching pattern. Initially, our malignant calcification tests returned benign results due to the ambiguous terminology. MammoNet node CalcificationShape uses the term linear branching while Tabar uses casting. Once we made the correct correlation, our results improved significantly.

8 Related Work

Relatively little work has been done in the development of decision-support systems for the diagnosis of breast disease. Heathfield et al. [13] describe a rule-based expert system for assisting pathologists in diagnosing fine needle aspirates of the breast. The system reaches five possible

diagnostic conclusions: malignant, suspicious of malignancy, indeterminate, probably benign, and benign. As the authors note, a fundamental inadequacy of their system is its inability to reason about the degrees of importance of diagnostic features, such as could be done using probabilities.

In more recent work, Heathfield et al. [14] describe a knowledge-based system for assisting pathologists in the histological diagnosis of breast disease. The system represents uncertainty using a qualitative measure consisting of five levels of certainty ranging between always and never.

D'orsi et al. [7] developed a reading aid and a computer-assisted decision aid to improve the accuracy of mammographic interpretation. The reading aid is a checklist of twelve features determined to have particular diagnostic value by studying the features used by radiologists expert in the interpretation of mammograms. Given a quantitative assessment of the twelve features, the decision aid reports the probability of malignancy.

9 Future Research

The emphasis in the first phase of MammoNet was its construction. Thorough testing of clinical cases is the focus of MammoNet's next phase, along with the associated refinement of the network. Further validation can be accomplished by measuring MammoNet's performance run on a test suite against that of a group of radiologists with differing levels of mammographic experience. Less experienced mammographers can use this network as an aid in diagnosis.

There are several potential enhancements for MammoNet. MammoNet used only the four commonly cited demographic risk factors, age, age of menarche, age of first live birth, and number of first degree relatives with a history of breast cancer. Recent epidemiological studies have focused on other demographic factors as possible causes of breast cancer. These factors: race, geographic location, hormone therapy, previous cancers, body build, and diet could be integrated into the demographic class of MammoNet.

The node *BreastCancer* is partitioned into two states: present or absent. *BreastCancer* could be refined to represent the numerous types of cancer and benign conditions, making reasoning about the disease more precise. Invasive lobular carcinoma presents frequently as asymmetric density and architectural distortion. Invasive ductal carcinoma presents more frequently as a spiculated mass [8]. Potential benign findings include cyst, fibroadenoma, and scar.

MammoNet's implementation, testing, and refinement phases could be automated through a mechanism which would update the link matrices with information stored in an associated database. MammoNet's decision-making ability could be extended from reasoning only about the

existence of the disease to reasoning about possible actions dependent on the mammographic findings. Biopsy is recommended when the probability of malignancy is high. Ultrasound is used when differentiating between solid masses and cysts. In other instances, the strategy is to wait a specified time period and to retest, looking for changes.

Acknowledgments

We thank Robert Krieger and Katherine A. Shaffer M.D. for many helpful discussions, and Olvi Mangasarian for the breast cancer information. A heartfelt "thank you" to Kun Wang who diligently helped calculate probabilities, enter data, and test the network. Dr. Haddawy was partially supported by NSF grant IRI-920762. This work was facilitated by a grant from the National Institutes of Health (USPHS G08 LM05705) to the Medical College of Wisconsin for Integrated Advanced Information Management Systems (IAIMS) planning.

References

1. Albertyn LE. Mammographically indeterminate microcalcifications-can we do better? Australsian Radiology 1991; 35: 350-357.

2. Bassett LW. Mammographic analysis of calcifications. Radiologic Clinics of North America 1992; 30: 93-105.

3. Ciatto S, et al. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. Radiology 1987; 165: 99-102.

4. Colditz GA. Family history, age, and risk of breast cancer prospective data from the Nurses' Health Study. JAMA 1993; 270: 338-343.

5. Díez FH. Parameter adjustment in Bayes networks. The generalized noisy-or gate. UAI-93 1993; 99-105.

6. Donegan WL. Evaluation of a palpable breast mass. Current Concepts 1992; 327: 937-942.

7. D'Orsi C, et al. Reading and decision aide for improved accuracy and standardization of mammographic diagnosis. Radiology 1992; 184: 619-622.

8. Egan RL, et al. Intramammary calcifications without an associated mass in benign and malignant diseases. Radiology 1980; 137: 1-7.

9. Feig SA. Breast masses, mammographic and sonographic evaluation. Radiologic Clinics of North America 1992; 30: 67-92.

10. Gail MH, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. ARTICLES 1989; 81: 1879-1886.

11. Haddawy P. Generating (Bayesian) networks from probability logic knowledge bases, Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, July 1994, 262-269.

12. Haddawy P, Kahn CE Jr, Butarbutar M. A Bayesian network model for radiological diagnosis and procedure selection: work-up of suspected gallbladder disease. Medical Physics 1994; 21: 1185-1192.

13. Heathfield HA, et al. Computer assisted diagnosis of fine needle aspirate. Journal of Clinical Pathology 1990; 43: 168-170.

14. Heathfield HA, et al. Knowledge-based computer system to aid in the histopathological diagnosis of breast cancer. Journal of Clinical Pathology 1991; 44; 502-508.

15. Jackson VP, et al. Diagnostic importance of the radiographic density of noncalcified breast masses: analysis of 91 lesions. AJR 1991; 157: 25-28.

16. Lay SF, et al. Breast biopsy changing patterns during a five-year period. American Surgeon 1990; 56: 79-85.

17. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine 1978; 4: 283-298.

18. Muir BB, et al. Microcalcification and its relationship to cancer of the breast: experience in a screening clinic. Clinical Radiology 1983; 34: 193-200.

19. Sickles EA. Breast calcifications: mammographic evaluation. Radiology 1986; 160: 289-293.

20. Sickles EA. Mammographic features of 300 consecutive nonpalpable breast cancers. American Journal of Roentgenology 1986; 146: 661-663.

21. Sickles EA. Mammograph of the postsurgical breast. American Journal of Roentgenology 1981, 136: 584-588

22. Srinivas S, Breese J. IDEAL: a software package for analysis of influence diagrams. Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence 1990; 212-219.

23. Symmonds RE, et al. Management of nonpalpable breast abnormalities. Ann. Surg 1987; 205: 520-528.

24. Swann CA, et al. The halo sign and malignant breast lesions. AJR 1987; 149: 1145-1147.

25. Tabar L, Dean PB. Teaching Atlas of Mammography. 2nd ed. New York, NY: Theime-Stratton, 1985.

26. Wu Y, et al. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. Radiology 1993; 81-87.